ELSEVIER Linear Algebra and its Applications 331 (2001) 43–59

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Hyperspheres and hyperplanes fitted seamlessly by algebraic constrained total least-squares☆

Yves Nievergelt

Department of Mathematics, Eastern Washington University, 216 Kingston Hall, Cheney, WA 99004-2418, USA

Received 22 December 1999; accepted 11 January 2001

Submitted by G.H. Golub

Abstract

For each finite set of points in a Euclidean space of any dimension, the algorithm presented here determines all the algebraically best fitting circles or lines, spheres or planes, or hyperspheres or hyperplanes, in a seamless manner from spherical through affine manifolds. In particular, affine submanifolds of any dimensions are *not* singularities of the algorithm. To this end, the algorithm combines projective geometry, Coope's and Gander et al.'s layouts of the equations, and Golub et al.'s generalization of the Schmidt–Mirsky matrix approximation theorem to solve the equations. The resulting best fitting manifolds remain invariant under rigid transformations. Moreover, if the best fitting manifold is affine, then it coincides with Golub and Van Loan's affine manifold of Total Least-Squares. Thus the algorithm can also fit hyperspheres in a manner that remains robust with data lying near a hyperplane. Furthermore, an analysis with a theorem of Wedin's shows that the fitted hyperspheres' sensitivity to perturbations of the data increases as the colinearity of the data increases. © 2001 Elsevier Science Inc. All rights reserved.

Keywords: Fitting; Circles; Spheres; Hyperspheres; Algebraic; Total least-squares

1. Introduction

The problem of fitting a circle or a sphere to observations arises in several applications. In computer science, partially automated graphic procedures involve fitting circles to points, for instance, to design fonts for typography [2,19]. In electrical

☆ This work was done at the University of Washington in Seattle during a leave from Eastern Washington University.

E-mail address: ynievergelt@ewu.edu (Y. Nievergelt).

PII: S 0 0 2 4 - 3 7 9 5 (0 1) 0 0 2 6 3 - 4

engineering, the calibration of microwave devices can proceed by fitting a circle in the complex plane to measurements of phase shifts caused by sliding the termination of a transmission line [12]. In particle physics, the identification of particles can include fitting circular trajectories to a large number of automated measurements of positions of electrically charged particles moving within a uniform magnetic field [5,16].

For such applications, the data consist of N points $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_N$ in the real Euclidean space $\mathbb{R}^L$, and the problem consists of determining a hypersphere

$$\mathcal{S}(\vec{\mathbf{c}}, r) := \{\vec{\mathbf{x}} \in \mathbb{R}^L : \|\vec{\mathbf{x}} - \vec{\mathbf{c}}\| = r\} \tag{1}$$

with center $\vec{\mathbf{c}} \in \mathbb{R}^L$ and radius r that minimize some objective.

With *affine* manifolds (hyperspheres with centers at infinity), minimizing the average squared distance from the data corresponds to the method of Total Least-Squares (TLS), initiated by Golub and Van Loan's Singular-Value Decomposition (SVD) of the data [9] and explained in detail by Van Huffel [25].

Attempts at adapting TLS to circles and spheres have been made by computer scientists [2,4,19], engineers [12], mathematicians [1,3,7,20,22], and physicists [5,16], trying to find "the" center $\vec{\mathbf{c}}$ and the radius r that minimize (with the Euclidean norm $\| \ \|_2$)

$$\sum_{k=1}^{N} \left(\|\vec{\mathbf{x}}_k - \vec{\mathbf{c}}\|_2 - r \right)^2, \tag{2}$$

but such attempts have led only to equations deemed "intractable" [5,20].

To circumvent such difficulties, computer scientists [2,19], engineers [12], mathematicians [10, pp. 191–196], and physicists [5,16], have substituted for the geometric distance an algebraic objective, in effect the total least-squares value of a function defining the fitted circle as its zero set, for example,

$$\sum_{k=1}^{N} \left(\|\vec{\mathbf{x}}_k - \vec{\mathbf{c}}\|_2^2 - r^2 \right)^2, \tag{3}$$

which leads to a linear system [5,12], or to ordinary least-squares [3]. Such algebraic methods are called *algebraic fits.* Because they minimize sums of squares, they can also be called methods of *algebraic total least-squares.* In contrast, methods that minimize geometric objectives, such as total least-squared distance, are *geometric fits*, and can also be called methods of *geometric total least-squares*.

More recently, however, Gander et al. have produced examples of data for which algebraic fits yield unsatisfactory results, in the form of circles lying far away from circles obtained from geometric fits [7, p. 561]. They have also argued that the results should remain invariant under rigid transformations of the space containing the data [7, p. 566]. They have further conjectured and illustrated by examples that algebraic fits could be the initial estimates for Newton's method to converge to the geometric fit. However, such geometric objectives "are prohibitively expensive compared to

the simple algebraic solution" [7, p. 577]. Moreover, there does not seem to exist any theorem guaranteeing the existence and uniqueness of a global minimum of the geometric objective. Indeed, for some data the geometric objective can have multiple global minima or no global minimum. Finally, there does not seem to exist any theorem guaranteeing the convergence of any algorithm toward any global minimum of any geometric objective from any initial estimate.

For applications where the data lie "very close" to a circle, there seems to be exactly one geometrically best fitting circle, and it appears to lie very close to the algebraically best fitting circle [5,12]. Consequently, the algebraically best fitting circle can serve as the final solution if it is satisfactory to the users [5,12], or it can serve as the initial estimate for a few refinements toward the geometrically best fitting circle [5, p. 224].

Nevertheless, existing algorithms to fit circles algebraically hitherto suffered from unrecoverable singularities if a point, line, or affine manifold fitted the data better than a circle, sphere, or hypersphere [19, Section 3, p. 147]. With floating-point computations, such algorithms can encounter unrecoverable singularities, undetected errors, or unsatisfactory results with points not necessarily on but merely near an affine manifold [19, Section 7, p. 149]. Therefore, the algorithm presented here offers for such applications the advantage of determining the best fitting curve or surface seamlessly, from circles through lines and points, or from spheres through planes, lines, and points, without encountering any unrecoverable singularity, and without involving any logical switch between spherical or affine manifolds. Finally, while the concept of a "satisfactory fit" remains subjective, the results presented below with test data from the literature suggest that the algorithm presented here produces fits that are more satisfactory than fits from other algebraic methods.

2. Previous algebraic methods

For each data point $\vec{\mathbf{x}}_i$ from any finite data set $\{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_N\} \subset \mathbb{R}^L$, define a function $f_i : \mathbb{R}^{L+1} \to \mathbb{R}$ by

$$\begin{aligned} f_i(\vec{\mathbf{c}}, r) &:= \|\vec{\mathbf{c}} - \vec{\mathbf{x}}_i\|_2^2 - r^2 \\ &= (\|\vec{\mathbf{c}}\|_2^2 - r^2) - \langle (2\vec{\mathbf{c}}), \vec{\mathbf{x}}_i \rangle + \|\vec{\mathbf{x}}_i\|_2^2, \end{aligned} \tag{4}$$

where $\langle\ ,\ \rangle$ is the Euclidean inner product. Thus the point $\vec{\mathbf{x}}_i$ lies on the hypersphere $\mathscr{S}(\vec{\mathbf{c}}, r)$ if and only if $f_i(\vec{\mathbf{c}}, r) = 0$. Consequently, all the data points lie on a common hypersphere $\mathscr{S}(\vec{\mathbf{c}}, r)$ if and only if $(\vec{\mathbf{c}}, r)$ is a zero of the function $F : \mathbb{R}^{L+1} \to \mathbb{R}$ defined by

$$F(\vec{\mathbf{c}}, r) := \sum_{i=1}^{N} \left[f_i(\vec{\mathbf{c}}, r) \right]^2 . \tag{5}$$

Calculus [16] shows that if F has a stationary point at $(\vec{\mathbf{c}}, r)$ then

$$r^2 = [R(\vec{\mathbf{c}})]^2 := \frac{1}{N}\sum_{i=1}^{N} \|\vec{\mathbf{c}} - \vec{\mathbf{x}}_i\|_2^2. \tag{6}$$

Thus minimizing F amounts to minimizing the function G defined by

$$G(\vec{\mathbf{c}}) := F(\vec{\mathbf{c}}, R(\vec{\mathbf{c}})) = \sum_{i=1}^{N} \{\|\vec{\mathbf{c}} - \vec{\mathbf{x}}_i\|_2^2 - [R(\vec{\mathbf{c}})]^2\}^2. \tag{7}$$

Because there are no constraints on $\vec{\mathbf{c}}$ and r, methods that minimize such functions as F or G can be called methods of *unconstrained algebraic total least-squares.* For the function G, calculus shows that the Hessian matrix of G equals $8N$ times the covariance of the data [17]. It follows that G has a unique global minimum if and only if there does not exist any hyperplane containing all the data points. In such a situation, the solution remains invariant under rigid Euclidean transformations, because so does the objective function F.

With Coope's change of coordinates [3, p. 384],

$$\vec{\mathbf{z}} := 2\vec{\mathbf{c}}, \quad z_0 := r^2 - \|\vec{\mathbf{c}}\|_2^2, \tag{8}$$

which admits an inverse defined for $z_0 + \|\vec{\mathbf{c}}\|_2^2 \geqslant 0$ by

$$\vec{\mathbf{c}} := \frac{1}{2}\vec{\mathbf{z}}, \quad r := \sqrt{z_0 + \|\vec{\mathbf{c}}\|_2^2}, \tag{9}$$

the equation $f_i(\vec{\mathbf{c}}, r) = 0$ becomes the linear equation $1 \cdot z_0 + \langle \vec{\mathbf{z}}, \vec{\mathbf{x}}_i \rangle = \|\vec{\mathbf{x}}_i\|_2^2$. Consequently, all the data points lie on a common hypersphere $\mathscr{S}(\vec{\mathbf{c}}, r)$ if and only if z_0 and $\vec{\mathbf{z}}$ satisfy the linear system

$$\begin{pmatrix} 1 & \vec{\mathbf{x}}_1^{\mathrm{T}} \\ \vdots & \vdots \\ 1 & \vec{\mathbf{x}}_N^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} z_0 \\ \vdots \\ z_L \end{pmatrix} = \begin{pmatrix} \|\vec{\mathbf{x}}_1\|_2^2 \\ \vdots \\ \|\vec{\mathbf{x}}_N\|_2^2 \end{pmatrix}. \tag{10}$$

For data that need not all lie on a common hypersphere, Coope's method consists in solving this linear system in the sense of ordinary least squares [3, p. 385]. The system admits a unique solution if and only if the rows $(1, \vec{\mathbf{x}}_1^{\mathrm{T}}), \ldots, (1, \vec{\mathbf{x}}_N^{\mathrm{T}})$ are linearly independent, which occurs if and only if there does not exist any hyperplane containing all the data points. In such a situation, the solution remains invariant under rigid Euclidean transformations, because so does the objective function F. However, if all the data points lie on a common hyperplane, then Coope's method fails, and so do other similar methods of Bookstein [2], Crawford [5], and Kåsa [12]. Moreover, in such cases of failure, such a remedy as the selection of the shortest least-squares solution (describe in [13, Chapter 14]) depends on the coordinate system and thus does *not* remain invariant under rigid Euclidean transformations.

Gander et al. [7] avoid the failures associated to hyperplanes by solving not for the center and radius but for the coefficients a, $\vec{\mathbf{b}}$, and c that minimize a function $F : \mathbb{R}^{L+2} \to \mathbb{R}$ defined by

$$F(a, \vec{\mathbf{b}}, c) := \sum_{i=1}^{N} \left(a\|\vec{\mathbf{x}}_i\|_2^2 + \langle \vec{\mathbf{b}}, \vec{\mathbf{x}}_i \rangle + c \right)^2, \tag{11}$$

subject to the constraint $\|(a, \vec{\mathbf{b}}^{\mathrm{T}}, c)\|_2 = 1$. Their solution consists in finding a right-singular vector $(a, \vec{\mathbf{b}}^{\mathrm{T}}, c) := \vec{\mathbf{v}}_{L+2}$ for the smallest singular value σ_{L+2} of the matrix

$$B := \begin{pmatrix} \|\vec{\mathbf{x}}_1\|_2^2 & \vec{\mathbf{x}}_1^{\mathrm{T}} & 1 \\ \vdots & \vdots & \vdots \\ \|\vec{\mathbf{x}}_N\|_2^2 & \vec{\mathbf{x}}_N^{\mathrm{T}} & 1 \end{pmatrix} = \sum_{j=1}^{L+2} \vec{\mathbf{u}}_j \sigma_j \vec{\mathbf{v}}_j^{\mathrm{T}}. \tag{12}$$

Hence $(a, \vec{\mathbf{b}}^{\mathrm{T}}, c)$ solves the linear system $(B - \sigma_{L+2}\vec{\mathbf{u}}_{L+2}\vec{\mathbf{v}}_{L+2}^{\mathrm{T}})(a, \vec{\mathbf{b}}^{\mathrm{T}}, c)^{\mathrm{T}} = \vec{\mathbf{0}}$. By the Schmidt–Mirsky theorem on matrix approximation [15,21], the matrix

$$\widehat{B} := B - \sigma_{L+2}\vec{\mathbf{u}}_{L+2}\vec{\mathbf{v}}_{L+2}^{\mathrm{T}} \tag{13}$$

is the matrix of rank $L+1$ closest to B relative to the Frobenius norm

$$\|B - \widehat{B}\|_F^2 = \sum_{i=1}^{N} \sum_{j=1}^{L+2} (B - \widehat{B})_{i,j}^2. \tag{14}$$

Algebraically, such a solution amounts to making in the matrix B the adjustments with the smallest sum of squares so that the system $\widehat{B}\vec{\mathbf{v}} = \vec{\mathbf{0}}$ has a non-zero solution $\vec{\mathbf{v}}_{L+1}$. The adjustments (the entries of $B - \widehat{B}$) are not quite adjustments of data in the sense of Deming [6], because they can also affect the constant column, so that $\widehat{B}$ can fail to have $\vec{\mathbf{1}}$ in its last column. In other words, in two different rows with indices $k \neq \ell$, whereas $B_{k,L+2} = 1 = B_{\ell,L+2}$ the adjusted entries can differ from each other, so that $\widehat{B}(k, L+2) \neq \widehat{B}(\ell, L+2)$. Consequently, the adjusted data points in the rows of $\widehat{B}$ satisfy different equations, with different constant terms $\widehat{B}(k, L+2)(\vec{\mathbf{v}}_{L+2})_{L+2} \neq \widehat{B}(\ell, L+2)(\vec{\mathbf{v}}_{L+2})_{L+2}$. This situation need not necessarily be detrimental, but it points to one direction for potential improvements. The same column $\vec{\mathbf{1}}$ also has a geometric significance. If the best fitting hypersurface is a hyperplane, so that $a = 0$, then such a solution minimizes

$$F(0, \vec{\mathbf{b}}, c) = \sum_{i=1}^{N} \left(\langle \vec{\mathbf{b}}, \vec{\mathbf{x}}_i \rangle + c \right)^2 = \sum_{i=1}^{N} \left[\left\langle (\vec{\mathbf{b}}^{\mathrm{T}}, c)^{\mathrm{T}}, (\vec{\mathbf{x}}_i^{\mathrm{T}}, 1)^{\mathrm{T}} \right\rangle \right]^2, \tag{15}$$

which is the sum of the squared distances from the "projectivized" data $(\vec{\mathbf{x}}_1^{\mathrm{T}}, 1), \ldots, (\vec{\mathbf{x}}_N^{\mathrm{T}}, 1)$ to the *projective* hyperplane perpendicular to the unit vector $(\vec{\mathbf{b}}^{\mathrm{T}}, c)$ and constrained to pass through the origin in $\mathbb{R}^{L+1}$. This hyperplane is *not* the TLS hyperplane in $\mathbb{R}^L$ that minimizes the sum of the squared Euclidean distances to the data. This explains why such algebraic fits fail to appear satisfactory for data lying near a hyperplane. Therefore, another method of solution becomes necessary. The foregoing algebraic and geometric considerations suggest that improvements may arise from excluding the constant column $\vec{\mathbf{1}}$ from any adjustments.

3. Definition of the algorithm

To accommodate hyperplanes and other affine submanifolds seamlessly, and to produce more satisfactory algebraic fits, the algorithm developed here modifies Gander et al.'s method in two ways.

Firstly, a translation shifts the center of coordinates to the mean of the data.

Secondly, the solution of the resulting overdetermined or underdetermined system utilizes not the Schmidt–Mirsky matrix approximation but, instead, Golub et al.'s generalization of such approximations by Constrained Total Least-Squares (CTLS), which constrains the first column (or any other submatrix) to remain constant. Because the objective function is algebraic, such methods of solution can also be called methods of *algebraic constrained total least-squares.*

The result produced by the algorithm remains invariant under rigid transformations, it keeps the column $\vec{\mathbf{1}}$ constant, and it forces the best fitting affine manifolds to coincide with the manifolds fitted by TLS. For affine manifolds, algebraic and geometric total least-squares coincide, because the sum of the squared distances happens to be a rational function.

Algorithm 1 (*Data*). Positive integers $L, N \in \mathbb{N}^*$, and a finite set $\{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_N\} \subset \mathbb{R}^L$.

Step 1. Compute the mean of the data,

$$\overline{\vec{\mathbf{x}}} := (1/N) \sum_{i=1}^{N} \vec{\mathbf{x}}_i, \tag{16}$$

and form the following matrix, with columns ordered as by Golub et al. [8]:

$$A := \begin{pmatrix} 1 & (\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}})^{\mathrm{T}} & \|\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}}\|_2^2 \\ \vdots & \vdots & \vdots \\ 1 & (\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}})^{\mathrm{T}} & \|\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}}\|_2^2 \end{pmatrix}. \tag{17}$$

Step 2. Apply Golub et al.'s method to find the matrix $\widehat{A}$ with rank $L+1$, with its first column $\vec{\mathbf{1}}$ kept constant, and closest to A relative to the Frobenius norm (minimizing $\|A - \widehat{A}\|_F$).

Step 3. Apply Golub et al.'s method to find the vector $\vec{\mathbf{v}}$ such that $\widehat{A}\vec{\mathbf{v}} = \vec{\mathbf{0}}$.

Merely to keep the first column $\vec{\mathbf{1}}$ constant, Golub et al.'s method simplifies as follows. The matrix A splits in the form $A = [A_1; A_2]$, with the first colum $A_1 := \vec{\mathbf{1}} = A(:, 1)$, and with $A_2 = (A(:, 2), \ldots, A(:, L+2))$ consisting of the remaining $L+1$ columns of A. Let Q be the orthogonal projection onto the column space of A_1, which is here

$$Q := (1/N)\vec{\mathbf{1}}\,\vec{\mathbf{1}}^{\mathrm{T}}. \tag{18}$$

Multiplication of any matrix $T \in \mathbb{M}_{N\times K}$ on the left by Q replaces every entry in each column $T(:, j)$ by the average $\overline{T(:, j)}$ of this column. For example, with

$$\text{mean}\left(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2\right) := \frac{1}{N}\sum_{j=1}^{N}\|\vec{\mathbf{x}}_j-\overline{\vec{\mathbf{x}}}\|_2^2, \tag{19}$$

and thanks to the shift to the mean $\overline{\vec{\mathbf{x}}}$, it follows that

$$QA_2 = \begin{pmatrix} \vec{\mathbf{0}}^{\mathrm{T}} & \text{mean}\,(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2) \\ \vdots & \vdots \\ \vec{\mathbf{0}}^{\mathrm{T}} & \text{mean}\,(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2) \end{pmatrix}. \tag{20}$$

Also, let $Q^{\perp} := I - Q$ be the orthogonal projection onto the orthogonal complement. Thus,

$$Q^{\perp}A_2 = \begin{pmatrix} (\vec{\mathbf{x}}_1-\overline{\vec{\mathbf{x}}})^{\mathrm{T}} & \|\vec{\mathbf{x}}_1-\overline{\vec{\mathbf{x}}}\|_2^2 - \text{mean}\,(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2) \\ \vdots & \vdots \\ (\vec{\mathbf{x}}_N-\overline{\vec{\mathbf{x}}})^{\mathrm{T}} & \|\vec{\mathbf{x}}_N-\overline{\vec{\mathbf{x}}}\|_2^2 - \text{mean}\,(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2) \end{pmatrix}. \tag{21}$$

Let $P := \min\{L+1, N\}$, and consider the Singular-Value Decomposition of $Q^{\perp}A_2$:

$$Q^{\perp}A_2 = \sum_{j=1}^{P}\sigma_j\vec{\mathbf{u}}_j\vec{\mathbf{v}}_j^{\mathrm{T}}, \tag{22}$$

with singular values ordered in non-increasing order $\sigma_1 \geqslant \cdots \geqslant \sigma_P \geqslant 0$. By the Schmidt–Mirsky matrix approximation theorem, the singular matrix closest (relative to the Frobenius norm) to $Q^{\perp}A_2$ is then

$$\widetilde{Q^{\perp}A_2} := \sum_{j=1}^{P-1}\sigma_j\vec{\mathbf{u}}_j\vec{\mathbf{v}}_j^{\mathrm{T}}. \tag{23}$$

In particular, the right-singular vector $\vec{\mathbf{v}}_P = (v_1, \ldots, v_L; v_{L+1})$ is a unit vector in the null space of $\widetilde{Q^{\perp}A_2}$. Moreover,

$$(QA_2)\vec{\mathbf{v}}_P = Q(A_2\vec{\mathbf{v}}_P) \in \text{Range}\,(A_1) = \text{Span}\,\{\vec{\mathbf{1}}\}, \tag{24}$$

whence there exists a real v_0 such that $(QA_2)\vec{\mathbf{v}}_P = -v_0\vec{\mathbf{1}}$. From QA_2 it follows that

$$(QA_2)\vec{\mathbf{v}}_P = \text{mean}\left(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2\right)\cdot v_{L+1}\cdot\vec{\mathbf{1}}, \tag{25}$$

whence

$$v_0 = -\text{mean}\left(\|\vec{\mathbf{x}}-\overline{\vec{\mathbf{x}}}\|_2^2\right)\cdot v_{L+1}. \tag{26}$$

Consequently,

$$\left[\vec{\mathbf{1}};\ \widetilde{Q^{\perp}A_2} + QA_2\right]\begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_L \\ v_{L+1} \end{pmatrix} = \vec{\mathbf{0}}. \tag{27}$$

Golub et al.'s work [8] shows that the matrix $[\vec{\mathbf{1}};\ \widetilde{Q^{\perp}A_2 + QA_2}]$ is the singular matrix that minimizes the Frobenius norm $\| [\vec{\mathbf{1}}; A_2] - [\vec{\mathbf{1}};\ \widetilde{Q^{\perp}A_2 + QA_2}] \|_F$.

Proposition 2. *The result of the algorithm remains invariant under rigid transformations.*

Proof. The result of the algorithm remains invariant under translations, because the initial subtraction of the mean annihilates every translation.

Moreover, each orthogonal matrix $U \in \mathbb{M}_{L\times L}(\mathbb{R})$ preserves the Euclidean norm, whence $\|U(\vec{\mathbf{x}}_j - \overline{\vec{\mathbf{x}}})\|_2 = \|\vec{\mathbf{x}}_j - \overline{\vec{\mathbf{x}}}\|_2$ for every data point. With

$$\mathscr{U} := \begin{pmatrix} 1 & \vec{\mathbf{0}}^{\mathrm{T}} & 0 \\ \vec{\mathbf{0}} & U & \vec{\mathbf{0}} \\ 0 & \vec{\mathbf{0}}^{\mathrm{T}} & 1 \end{pmatrix} \tag{28}$$

it follows that a rotation by U transforms the matrix A into a new matrix, which factors in the form $A\mathscr{U}^{\mathrm{T}}$:

$$\begin{aligned} &\begin{pmatrix} 1 & [U(\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}})]^{\mathrm{T}} & \|U(\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}})\|_2 \\ \vdots & \vdots & \vdots \\ 1 & [U(\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}})]^{\mathrm{T}} & \|U(\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}})\|_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & [U(\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}})]^{\mathrm{T}} & \|\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}}\|_2 \\ \vdots & \vdots & \vdots \\ 1 & [U(\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}})]^{\mathrm{T}} & \|\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}}\|_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & (\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}})^{\mathrm{T}} & \|\vec{\mathbf{x}}_1 - \overline{\vec{\mathbf{x}}}\|_2 \\ \vdots & \vdots & \vdots \\ 1 & (\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}})^{\mathrm{T}} & \|\vec{\mathbf{x}}_N - \overline{\vec{\mathbf{x}}}\|_2 \end{pmatrix} \begin{pmatrix} 1 & \vec{\mathbf{0}}^{\mathrm{T}} & 0 \\ \vec{\mathbf{0}} & U^{\mathrm{T}} & \vec{\mathbf{0}} \\ 0 & \vec{\mathbf{0}}^{\mathrm{T}} & 1 \end{pmatrix}. \end{aligned} \tag{29}$$

Therefore, the system $A\vec{\mathbf{v}} = \vec{\mathbf{0}}$ is equivalent to the system $(A\mathscr{U}^{\mathrm{T}})(\mathscr{U}\vec{\mathbf{v}}) = \vec{\mathbf{0}}$, which means that a rotation of the space of the data by U also rotates the solution (center or normal vector) by U, without changing the radius because $\mathscr{U}_{1,1} = 1$. Since the orthogonal matrix $\mathscr{U}$ preserves the first column $\vec{\mathbf{1}}$, and preserves the Frobenius norm of any matrix on its left and the Euclidean norm of any vector on its right, it follows that it also preserves Golub et al.'s solution. □

4. Geometric significance of the algorithm

The geometric significance of the right-singular vectors produced by the algorithm depends on the last coordinates and on the singular values of the matrix.

Generically, $N > L$, $P = L + 1$, $\sigma_L > \sigma_{L+1}$, whence there exists exactly one (algebraically) best fitting hypersurface, and one of two alternatives occurs.

If $v_{L+1} \neq 0$, then the best fitting hypersurface is a hypersphere, and a division by $-v_{L+1}$ gives

$$(z_0; z_1, \ldots, z_N; -1) = -(1/v_{L+1}) \cdot (v_0; v_1, \ldots, v_N; v_{L+1}). \tag{30}$$

The radius $r = \sqrt{z_0 + \|\vec{\mathbf{z}}\|_2^2}$ is real, because $v_0 = -v_{L+1}$ mean $(\|\vec{\mathbf{x}} - \overline{\vec{\mathbf{x}}}\|_2^2)$ whence

$$z_0 = -v_0/v_{L+1} = \text{mean}\,(\|\vec{\mathbf{x}} - \overline{\vec{\mathbf{x}}}\|_2^2) \geqslant 0. \tag{31}$$

The center is $\frac{1}{2}\vec{\mathbf{z}}$ relative to the mean $\overline{\vec{\mathbf{x}}}$, and hence $\vec{\mathbf{c}} = \overline{\vec{\mathbf{x}}} + \frac{1}{2}\vec{\mathbf{z}}$ relative to the initial coordinates.

If $v_{L+1} = 0$, then $v_0 = -\text{mean}\,(\|\vec{\mathbf{x}} - \overline{\vec{\mathbf{x}}}\|_2^2) \cdot v_{L+1} = 0$ also, and hence

$$\left[\vec{\mathbf{1}};\ \widetilde{Q^{\perp}A_2} + QA_2\right] \begin{pmatrix} 0 \\ v_1 \\ \vdots \\ v_L \\ 0 \end{pmatrix} = \left[\widetilde{Q^{\perp}A_2}(:,1), \ldots, \widetilde{Q^{\perp}A_2}(:,L)\right] \begin{pmatrix} v_1 \\ \vdots \\ v_L \end{pmatrix} = \vec{\mathbf{0}}. \tag{32}$$

Thus the unit vector $(v_1, \ldots, v_L)$ is perpendicular to every row of the L columns $[\widetilde{Q^{\perp}A_2}(:,1), \ldots, \widetilde{Q^{\perp}A_2}(:,L)]$. This means that there exists exactly one (algebraically) best fitting hypersurface among hyperspheres and hyperplanes, which is a hyperplane passing through the new center of coordinates $\overline{\vec{\mathbf{x}}}$ with unit normal vector $(v_1, \ldots, v_L)$. However, thanks to the shift of coordinates to the mean of the data, the first L columns of QA_2 equal zero, so that the first L columns of $\widetilde{Q^{\perp}A_2}$ coincide with the first L columns of A_2. Because of the minimizing property of $\widetilde{Q^{\perp}A_2}$, it follows that the first L columns of $\widetilde{Q^{\perp}A_2}$ form the singular matrix closest to the first L columns of A_2 relative to the Frobenius norm. Otherwise there would exist a still closer singular matrix. By the Schmidt–Mirsky matrix approximation theorem, it follows that the unit vector $(v_1, \ldots, v_L)$ is also the right-singular vector that minimizes the Euclidean norm $\|\,[A_2(:,1), \ldots, A_2(:,L)]\vec{\mathbf{v}}\,\|_2$. Therefore the fitted hyperplane coincides with the hyperplane of TLS, which minimizes the average squared Euclidean distance to the data [18].

In the special case $N \leqslant L$ or $\sigma_L = \sigma_{L+1}$, one of several possibilities can occur. In general, the equalities

$$\sigma_{L+1-K} > \sigma_{(L+1-K)+1} = \cdots = \sigma_{L+1} \tag{33}$$

mean that there exist a set of hypersurfaces, parametrized by the K-dimensional linear space of right-singular vectors corresponding to the mutually equal singular values $\sigma_{(L+1-K)+1} = \cdots = \sigma_{L+1}$, that (algebraically) fit the data equally best. Each hypersurface has a type given by the last coordinate of the corresponding sigular vector, $(\vec{\mathbf{v}}_{L-I})_{L+1}$.

For example, if $\sigma_{L-1} > \sigma_L = \sigma_{L+1}$ and $(\vec{\mathbf{v}}_L)_{L+1} = 0 = (\vec{\mathbf{v}}_{L+1})_{L+1}$, then the best fitting hypersurfaces are two non-parallel hyperplanes, which intersect in a best fitting straight line.

Similarly, if $\sigma_{L-1} > \sigma_L = \sigma_{L+1}$ but $(\vec{\mathbf{v}}_L)_{L+1} \neq 0 = (\vec{\mathbf{v}}_{L+1})_{L+1}$, then the best fitting hypersurfaces are a hypersphere and a hyperplane, which intersect in a best fitting circle (or a hypersphere of lower dimension) in space.

Finally, if $\sigma_1 > \sigma_2 = \cdots = \sigma_{L+1}$ and $(\vec{\mathbf{v}}_2)_{L+1} = \cdots = (\vec{\mathbf{v}}_{L+1})_{L+1} = 0$, then the best fitting hypersurfaces are L hyperplanes with linearly independent normals, which intersect in one best fitting point in space.

In particular, points and other affine submanifolds are *not* singularities of the algorithm.

5. Sensitivity to perturbations of the data

The result of the algorithm consists of at least one singular vector $\vec{\mathbf{v}}_{L+1}$ of a matrix $Q^\perp A_2$ with entries depending on the data. Consequently, perturbing the data from an initial set $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_N$ to a new set $\vec{\mathbf{x}}'_1, \ldots, \vec{\mathbf{x}}'_N$ also perturbs the matrix from A_2 to a new matrix A'_2. In either case Q still represents the orthogonal projection on $A_1 = \vec{\mathbf{1}}$, whence the perturbations change $Q^\perp A_2$ into $Q^\perp(A'_2)$. Because Q is an orthogonal matrix, with Euclidean norm $\|Q\|_2 = 1$, it follows that $\|Q^\perp(A'_2) - Q^\perp A_2\|_2 \leqslant \|A'_2 - A_2\|_2$. To analyze the induced perturbation on the singular vectors, denote the singular values and vectors of the initial matrix $Q^\perp A_2$ by $\sigma_1 \geqslant \cdots \geqslant \sigma_{L+1}$ and $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_{L+1}$, and those of the perturbed matrix $Q^\perp(A'_2)$ by $\sigma'_1 \geqslant \ldots \geqslant \sigma'_{L+1}$ and $\vec{\mathbf{v}}'_1, \ldots, \vec{\mathbf{v}}'_{L+1}$. With the notation $\Delta A_2 := A'_2 - A_2$ and $\Delta\vec{\mathbf{v}}_{L+1} = \vec{\mathbf{v}}'_{L+1} - \vec{\mathbf{v}}_{L+1}$, Van Huffel's version [25, p. 216, (7.45)] of Wedin's "generalized $\sin(\theta)$ theorem" [26, p. 102] gives an upper bound for the perturbation of the result, $\Delta\vec{\mathbf{v}}_{L+1}$, in terms of perturbation of the matrix, ΔA_2:

$$\|\Delta\vec{\mathbf{v}}_{L+1}\|_2 \leqslant \frac{\|Q^\perp \Delta A_2\|_2}{\sigma'_L - \sigma_{L+1}} \leqslant \frac{\|\Delta A_2\|_2}{\sigma'_L - \sigma_{L+1}}. \tag{34}$$

The inequalities just obtained (34) involve the gap between the smallest singular value σ_{L+1} of the initial matrix, $Q^\perp A_2$, and the next smallest singular value σ'_L of the perturbed matrix, $Q^\perp(A'_2)$. However, from theorems on the induced perturbations of singular values [24, p. 562], it follows that

$$|\sigma'_L - \sigma_L| \leqslant \|Q^\perp \Delta A_2\|_2. \tag{35}$$

If $\|Q^\perp \Delta A_2\|_2 < \sigma_L$, then a combination of inequalities (34) and (35) gives

$$\|\Delta\vec{\mathbf{v}}_{L+1}\|_2 \leqslant \frac{\|Q^\perp \Delta A_2\|_2}{(\sigma_L - \|Q^\perp \Delta A_2\|_2) - \sigma_{L+1}}. \tag{36}$$

Because singular vectors have unit Euclidean norms, and because $\|Q^\perp A_2\|_2 = \sigma_1$, inequality (36) can also take the following form:

$$\frac{\|\Delta\vec{\mathbf{v}}_{L+1}\|_2}{\|\vec{\mathbf{v}}_{L+1}\|_2} \leqslant \frac{\|Q^{\perp}\Delta A_2\|_2}{\|Q^{\perp}A_2\|_2} \frac{\sigma_1}{(\sigma_L - \|Q^{\perp}\Delta A_2\|_2) - \sigma_{L+1}}. \tag{37}$$

Consequently,

$$\limsup_{\|Q^{\perp}\Delta A_2\|_2 \to 0} \frac{\|\Delta\vec{\mathbf{v}}_{L+1}\|_2/\|\vec{\mathbf{v}}_{L+1}\|_2}{\|Q^{\perp}\Delta A_2\|_2/\|Q^{\perp}A_2\|_2} \leqslant \frac{\sigma_1}{\sigma_L - \sigma_{L+1}}. \tag{38}$$

Thus the factor $\kappa := \sigma_1/(\sigma_L - \sigma_{L+1})$ plays the role of the "condition number" [1, p. 38] for the problem of fitting generalized hyperspheres to data.

Geometrically, the gap $\sigma_L - \sigma_{L+1}$ is nearly zero if and only if the two corresponding generalized hyperspheres (with equations defined by the singular vectors $\vec{\mathbf{v}}_L$ and $\vec{\mathbf{v}}_{L+1}$) fit the data nearly equally well, as measured by the sum of squares σ_L^2 and σ_{L+1}^2. Because a linear combination of $\vec{\mathbf{v}}_L$ and $\vec{\mathbf{v}}_{L+1}$ can anihilate the last coordinate, their intersection, which is a hyperplane, also fits the data nearly as well as the two generalized hyperspheres do. In this sense the sensitivity of the results increases as the data becomes more colinear.

6. Applications and examples

Example 3. To investigate the design of an ancient Greek stadium, Rorres and Romano [20] fitted a circle to 21 points measured along the starting line for foot races. Their data consist of 21 pairs of Cartesian coordinates, listed in [20, Table 1, p. 748], and displayed here in Fig. 1. Their problem consists in fitting a circle to the data, without any requirements on the objective to be minimized. Their solution was to use an iterative method to fit the circle minimizing the sum of squared distances (the circle of geometric total least-squares).

Fig. 1 displays the circle fitted by the method of algebraic CTLS just presented (Algorithm 1). For this example, $N = 21$ and $L = 2$. The algorithm produced the $L + 1$ singular values

$$\sigma_1 = 49.674999, \quad \sigma_2 = 16.213949, \quad \sigma_3 = 0.059729973. \tag{39}$$

Because $\sigma_L > \sigma_{L+1}$, there exists only one best fitting curve. Moreover, the condition number takes the moderate value

$$\kappa = \frac{\sigma_1}{\sigma_2 - \sigma_3} \approx 3.075. \tag{40}$$

Furthermore, the corresponding right-singular vector is

Table 1
Comparison of methods for Rorres and Romano's data [20]

Source	Method	Center	Radius
Rorres and Romano [20]	Iterative	(−20.940, 33.618)	53.960
Algorithm 1	Algebraic	(−20.944, 33.616)	53.964

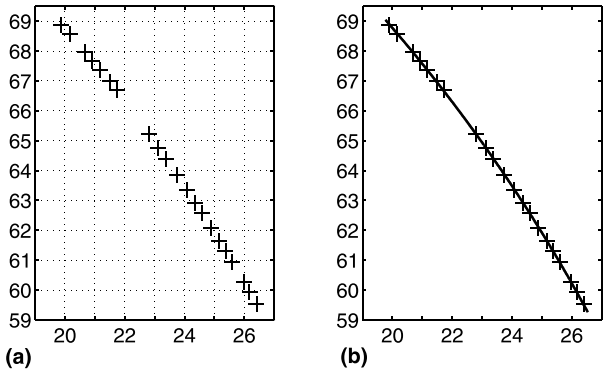

Fig. 1. (a) Data from Rorres and Romano [20]. (b) Fitted CTLS circle.

$$\begin{aligned}\vec{\mathbf{v}} &= (v_0;\, v_1, v_2, v_3)\\ &= (-0.12031091;\ 0.82381441,\ 0.56678355,\ 0.0092857035).\end{aligned} \tag{41}$$

The Singular-Value Decomposition of $\widetilde{QA_2}$ normalizes $\|(v_1, \ldots, v_{L+1})\|_2 = 1$ without taking v_0 into account, because Golub et al.'s method determines v_0 later. From $v_{L+1} = 0.0092857035 \neq 0$, it follows that the best fitting curve is a circle. This circle has radius 53.964077 and center at $(-20.943718,\ 33.615613)$, which on the scale of Fig. 1 is indistinguishable from Rorres and Romano's circle. Table 1 compares both circles.

Example 4. Mamakov [14] measured the coordinates of 32 craters on the Moon. His data consist of 32 triples of Cartesian coordinates, $(\xi_1, \eta_1, \zeta_1), \ldots, (\xi_{32}, \eta_{32}, \zeta_{32})$, dimensionless after scaling by the radius of the Moon, and displayed here in Fig. 2. Mamakov's objective was not necessarily to fit a surface, but to improve on previous systems of selenodetic coordinates by establishing a new one based on heliometric observations. To this end he took for granted that the surface of the Moon was modeled by a sphere of radius 1738 km [11, p. 476]. Therefore, in contrast to Example 3, which fitted a yet unknown curve, this example aims merely at testing the algorithms presented here against a known surface, namely a unit sphere.

Fig. 2 displays the result. For this example, $N = 32$ and $L = 3$. The algorithm produced the $L + 1$ singular values

$$\begin{aligned}\sigma_1 &= 1.4955523 \times 10^5,\\ \sigma_2 &= 2.8998535,\\ \sigma_3 &= 2.3743396,\\ \sigma_4 &= 0.012136124.\end{aligned} \tag{42}$$

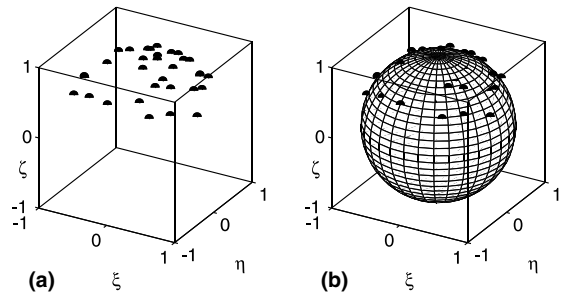

Fig. 2. (a) Thirty-two craters from Mamakov [14]. (b) Fitted CTLS sphere.

Because $\sigma_L > \sigma_{L+1}$, there exists only one best fitting surface, with

$$\kappa = \frac{\sigma_1}{\sigma_3 - \sigma_4} \approx 63,312. \tag{43}$$

Also, the corresponding right-singular vector is

$$\begin{aligned}\vec{\mathbf{v}} &= (v_0;\, v_1, v_2, v_3, v_4)\\ &= (-0.367521 \cdot 10^5;\ 0.107167,\ 0.0852749,\ 0.990577,\\ &\quad 0.715826 \times 10^{-5}).\end{aligned} \tag{44}$$

Because $v_{L+1} = 0.71582582 \times 10^{-5} \neq 0$, it follows that the best fitting surface is a sphere. The fitted CTLS sphere has radius 1.000 658 Moon radius.

Example 5. Gander et al. proposed the following data to demonstrate the inadequacy of algebraic fits and the superiority of geometric fits, in particular, least-squares fits [7, p. 560]:

$$(1,7), \quad (2,6), \quad (3,7), \quad (5,8), \quad (7,7), \quad (9,5). \tag{45}$$

This example shows that for these data the CTLS circle gives a better fit than their algebraic fit, perhaps as adequate as their least-squares fit.

Fig. 3 displays the result. For this example, $N = 6$ and $L = 2$. The algorithm produced the $L + 1$ singular values

$$\sigma_1 = 18.112127, \quad \sigma_2 = 5.9825553, \quad \sigma_3 = 1.3446374. \tag{46}$$

Again because $\sigma_L > \sigma_{L+1}$, there exists only one best fitting curve, with

$$\kappa = \frac{\sigma_1}{\sigma_2 - \sigma_3} \approx 3.905. \tag{47}$$

The corresponding right-singular vector is

$$\begin{aligned}\vec{\mathbf{v}} &= (v_0;\, v_1, v_2, v_3)\\ &= (-0.93528931;\ -0.0019315095,\ 0.99434122,\ 0.10621582).\end{aligned} \tag{48}$$

Because $v_{L+1} = 0.10621582 \neq 0$, it follows that the best fitting curve is a circle. This CTLS circle has radius 5.542 124 and center at (4.509 0924, 1.985 9081).

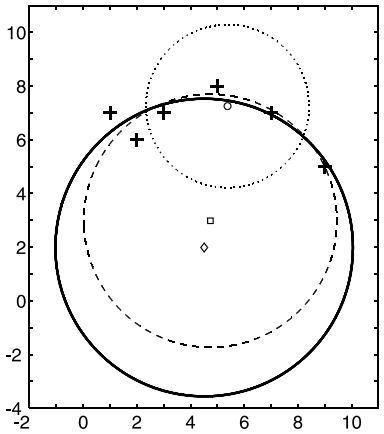

Fig. 3. Data (+) from Gander et al. [7]; fitted algebraic (· · ·), geometric (––), and CTLS (—) circles, with respective centers (○), (□), and (◊).

Table 2
Comparison of methods for Gander et al.'s data [7]

Source	Method	Center	Radius
Gander et al. [7]	Algebraic	(5.3794, 7.2532)	3.0370
Gander et al. [7]	Iterative	(4.7398, 2.9835)	4.7142
Algorithm 1	Algebraic	(4.5091, 1.9859)	5.5421

Table 2 compares this result with other circles fitted by other methods. While the circle fitted algebraically by Gander et al. exhibits the excessive curvature already observed by Pratt [19, Section 7, p. 149, Fig. 1], the CTLS circle fitted here remains closer to the geometric circle (also fitted by Gander et al.).

Example 6. Späth proposed the following data, also shown in Fig. 4, for comparisons of methods to fit circles [23, pp. 347–348, Example 2]:

$$(-1,-1),\ (-1,7),\ (0,3),\ (2,8),\ (5,-3),\ (7,5),\ (9,9),\ (10,3). \quad (49)$$

For this example, $N = 6$ and $L = 2$. The algorithm produced the $L + 1$ singular values

$$\sigma_1 = 42.570766, \quad \sigma_2 = 12.485514, \quad \sigma_3 = 9.7177986. \quad (50)$$

Because $\sigma_L > \sigma_{L+1}$, there exists only one best fitting curve, with

$$\kappa = \frac{\sigma_1}{\sigma_2 - \sigma_3} \approx 15.381. \quad (51)$$

Also, the corresponding right-singular vector is

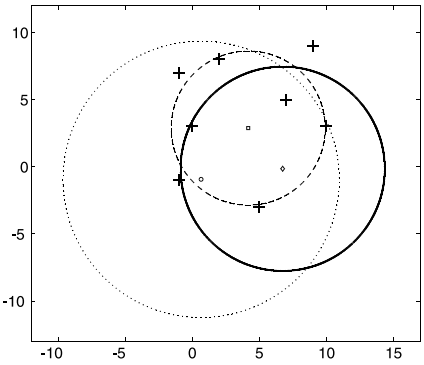

Fig. 4. Algebraic ($\cdots$), geometric ($--$), and CTLS ($—$) circles, with respective centers ($\circ$), ($\square$), and ($\Diamond$), fitted to Späth's data [23].

Table 3
Comparison of methods for Späth's data [23]

Source	Method	Center	Radius
Algorithm [7]	Algebraic	(0.6625, −0.9365)	10.291
Späth [23]	Iterative	(4.172, +2.879)	5.739
Algorithm 1	Algebraic	(6.739, −0.1497)	7.607

$$\begin{aligned}\vec{\mathbf{v}} &= (v_0, v_1, v_2, v_3)\\ &= (-3.3706460,\ -0.57679122,\ 0.81065981,\ 0.10071024).\end{aligned} \tag{52}$$

Because $v_{L+1} = 0.10071024 \neq 0$, it follows that the best fitting curve is a circle. Fig. 4 and Table 3 compare this circle with circles fitted by other methods: Gander et al.'s algebraic fit, and Späth's circle of geometric total least-squares fitted iteratively.

Example 7. To compare the circles fitted to increasingly colinear data by various methods, Pratt [19, Section 7, p. 149, Fig. 1] proposed data sets of the form

$$(-1, 0) \quad (-0.3, y) \quad (0.3, 0.1) \quad (1, 0), \tag{53}$$

with $y \in \{0.1, 0.02, -0.02, -0.06\}$. For $y = 0.1$, all the data points lie on a circle, which all the methods examined here reproduce exactly. For $y \neq 0.1$, Pratt shows that some algebraic methods produce circles of increasing curvature as the colinearity of the data increases. To test the invariance of the results under translations, the present examples translate the data by $\vec{\mathbf{s}} := (8, 9)$, which give

$$(7, 9) \quad (7.7, y) \quad (8.3, 9.1) \quad (9, 9) \tag{54}$$

with $y \in \{9.1, 9.02, 8.98, 8.94\}$. Because the result changes dramatically from $y = 9.1$ to $y = 9.02$, the present examples also contains three additional stages inserted between Pratt's first two stages, and one additional stage at the end with $y = 8.91$. Thus, $y \in \{9.1, 9.08, 9.06, 9.04, 9.02, 8.98, 8.94, 8.91\}$.

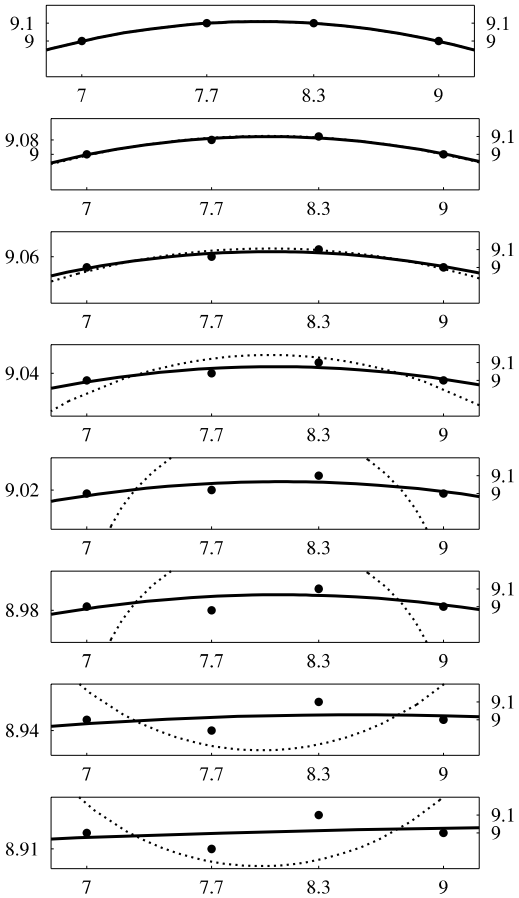

Fig. 5. Data (•) adapted from Pratt [19]; algebraic circle (· · ·) fitted by the algorithm from [7], and CTLS circle (—) fitted by Algorithm 1.

Fig. 5 shows that as the colinearity of the data increases, the circle fitted by Gander et al.'s algebraic method increases its curvature, while the CTLS circle fitted by the algorithm presented here decreases its curvature and approaches the line of total least-squares.

References

[1] Åke Björk, Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996.

[2] F.L. Bookstein, Fitting conic sections to scattered data, Comput. Graphics Image Process. 9 (1979) 56–71.
[3] I.D. Coope, Circle fitting by linear and nonlinear least squares, J. Optim. Theory Appl. 76 (2) (1993) 381–388.
[4] M.G. Cox, H.M. Jones, An algorithm for least-squares circle fitting to data with specified uncertainty ellipses, IMA J. Numer. Anal. 9 (3) (1989) 285–298.
[5] J.F. Crawford, A non-iterative method for fitting circular arcs to measured points, Nucl. Instr. and Meth. 211 (2) (1983) 223–225.
[6] W. Edwards Deming, Statistical Adjustment of Data, Dover, New York, 1964.
[7] W. Gander, G.H. Golub, R. Strebel, Least-squares fitting of circles and ellipses, BIT 34 (1994) 558–578.
[8] G.H. Golub, A. Hoffman, G.W. Stewart, A generalization of the Eckart–Young–Mirsky matrix approximation theorem, Linear Algebra Appl. 88&89 (1987) 317–327.
[9] G.H. Golub, C.F. Van Loan, An analysis of the total least squares problem, SIAM J. Numer. Anal. 17 (6) (1980) 883–893.
[10] P. Henrici, Essentials of Numerical Analysis with Pocket Calculator Demonstrations, Wiley, New York, 1982.
[11] H. Karttunen, P. Kröger, H. Oja, M. Poutanen, K. Johan Donner (Eds.), Fundamental Astronomy, second enlarged ed., Springer, New York, 1994.
[12] I. Kåsa, A circle fitting procedure and its error analysis, IEEE Trans. Instrum. Meas. 25 (1) (1976) 8–14.
[13] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Classics in Applied Mathematics, vol. 15, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995.
[14] A.S. Mamakov, System of selenodetic coordinates based on heliometric observations of the Moon at kazan, The Moon and the Planets 21 (1) (1979) 19–23.
[15] L. Mirsky, Symmetric gauge functions and unitarily invariant norms, Quart. J. Math. (Oxford) 11 (1960) 50–59.
[16] L. Moura, R. Kitney, A direct method for least-squares circle fitting, Comput. Phys. Commun. 64 (1) (1992) 57–63.
[17] Y. Nievergelt, Computing circles and spheres of arithmetic least squares, Comput. Phys. Commun. 81 (3) (1994) 343–350.
[18] Y. Nievergelt, Total least squares: state-of-the-art regression in numerical analysis, SIAM Rev. 36 (2) (1994) 258–264.
[19] V. Pratt, Direct least-squares fitting of algebraic surfaces, ACM Comput. Graphics 21 (4) (1987) 145–152.
[20] C. Rorres, D.G. Romano, Finding the center of a circular starting line in an ancient Greek stadium, SIAM Rev. 39 (4) (1997) 745–754.
[21] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen, 1. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener, Math. Ann. 63 (1907) 433–476.
[22] H. Späht, Least-squares fitting with spheres, J. Optim. Theory Appl. 96 (1) (1998) 191–199.
[23] H. Späth, Orthogonal distance fitting by circles and ellipses with given area, Comput. Statist. 12 (1997) 343–354.
[24] G.W. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (4) (1993) 551–566.
[25] S.V. Huffel, J. Vandewalle, The Total Least Squares Problem: Computational Aspects and Analysis, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1991.
[26] P.-Å. Wedin, Perturbation bounds in connection with singular value decomposition, BIT 12 (1) (1972) 99–111.